

LA-UR-21-27444

Approved for public release; distribution is unlimited.

Title: The Importance of Ontologies to LANL's Future Search Capabilities

Author(s): Chadwick, Thomas Anthony
Manginell, Sydney Joan
Mason, William Isaac MacGillivray
Valdez, Camille Danielle


Intended for: Symposium Presentation

Issued: 2021-07-29

Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.



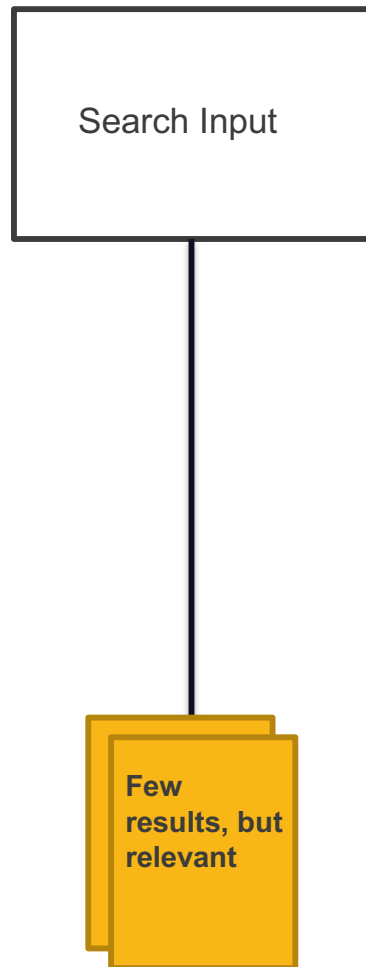
A black and white photograph of the Fat Man atomic bomb, a large, bulbous, egg-shaped weapon with a complex metal framework on top and four tail fins at the rear. It is mounted on a transport cradle with wheels. The bomb is positioned diagonally across the frame, from the bottom left towards the top right. The background is a light, flat surface.

The Importance of Ontologies to LANL's Future Search Capabilities

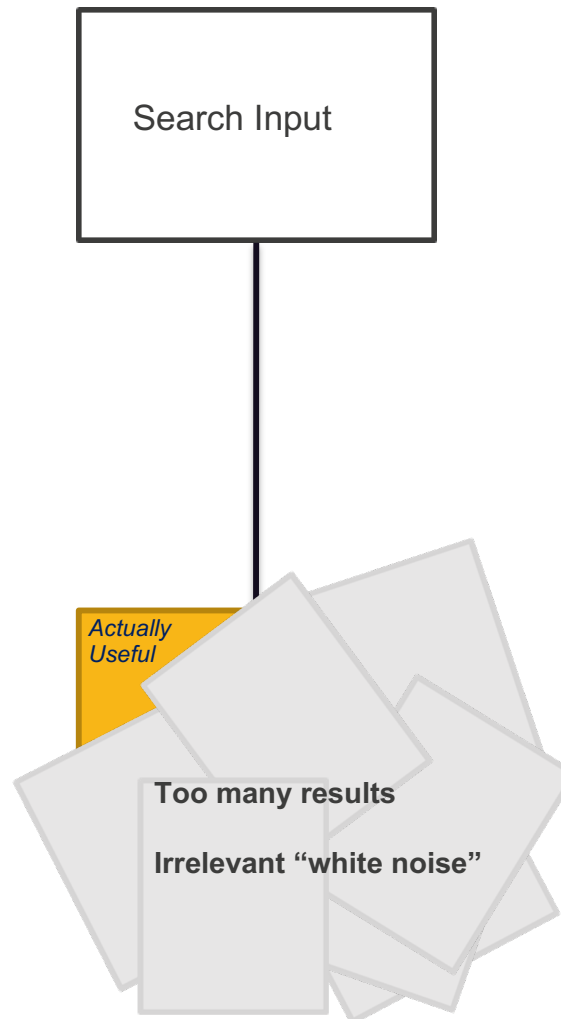
Thomas Chadwick, Sydney Manginell,
William Mason, and Camille Valdez

Search engines: recall, precision and the F-score golden mean

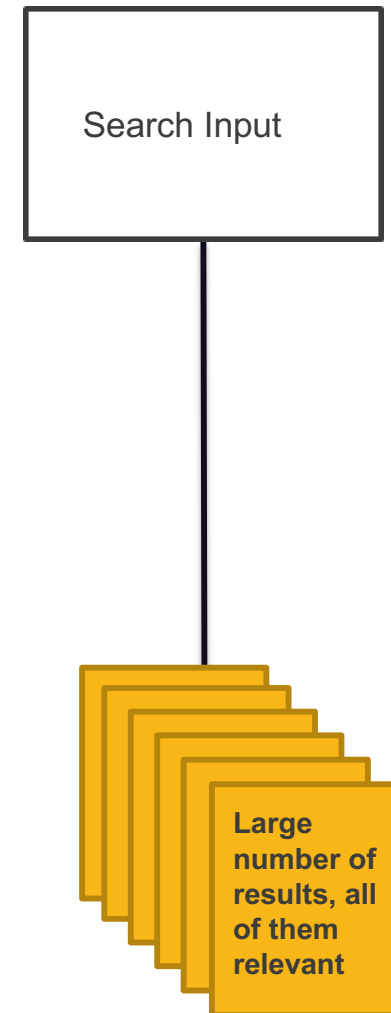
Low recall, high precision



High recall, low precision

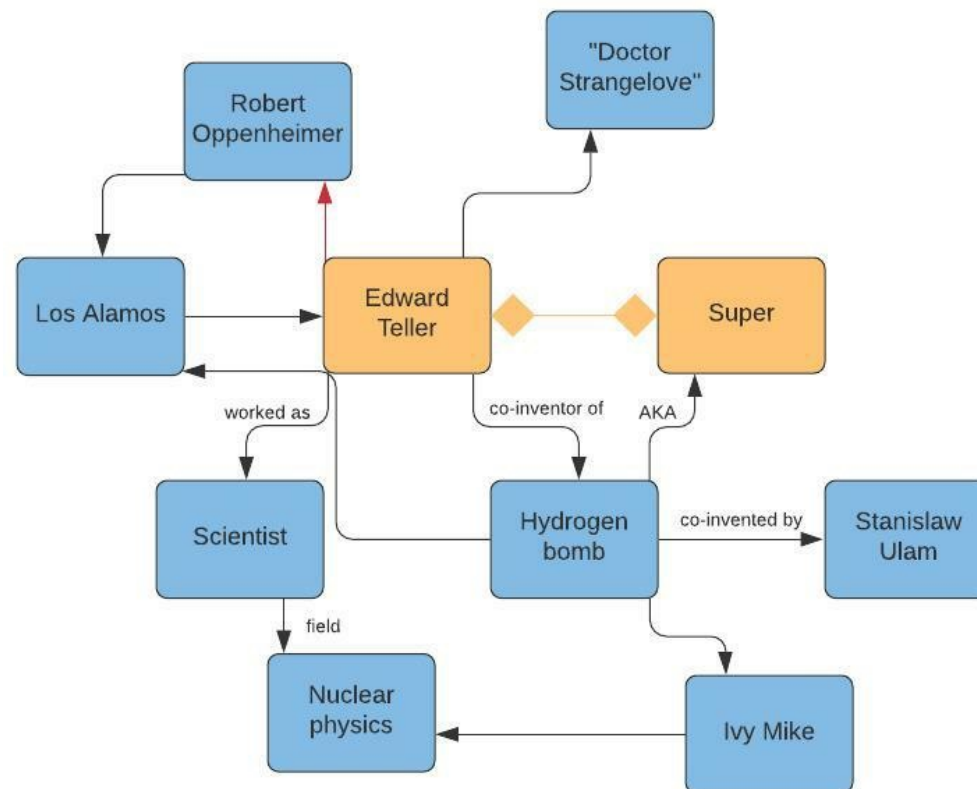


F-score: recall = precision

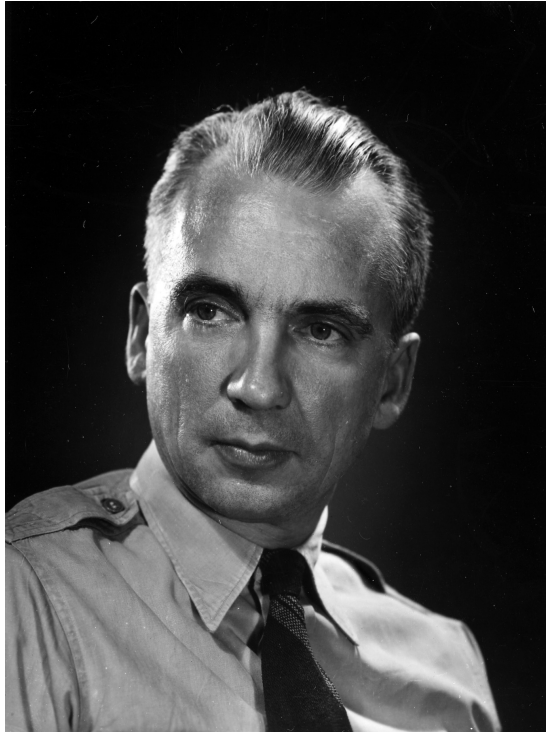


But what is an ontology?

- A network of concepts and the relationships joining them
- Ontologies make databases searchable
- Better recall from ability to link search terms with related concepts
- Better precision from ability to determine a document's relevancy
- Search string: "Edward Teller super"



Norris Bradbury and Edward Teller



Norris Bradbury was the second Director of the Los Alamos Scientific Laboratory (LASL). Bradbury rebuilt LASL after the postwar exodus and was instrumental in LASL's production of efficient fission bombs and the thermonuclear program.

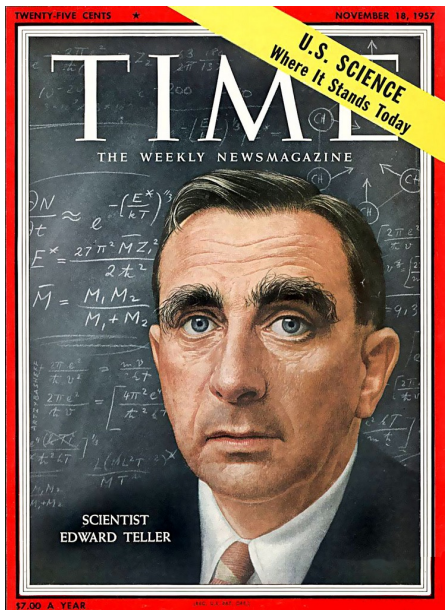


Edward Teller was a brilliant theoretical physicist who is rightly known as the key architect behind the Hydrogen bomb. Teller worked intermittently at Los Alamos before becoming the co-founder of the now-named Lawrence Livermore National Laboratory (LLNL).

A Tense Relationship over the H-Bomb

“The philosophy of Teller during the war has mystified many people. His fanatical obsession [with the Super] worked to the immediate detriment of this country’s objectives.” – Bradbury

The success of the H-Bomb is due “to the persistent and superhuman efforts of LASL” rather than “the dubious paternity of Teller.” – Bradbury

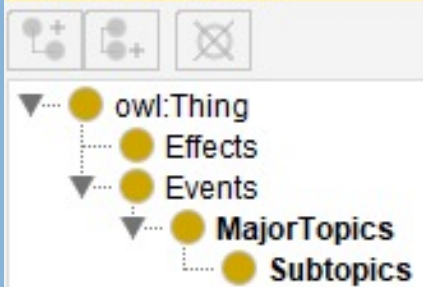


“I would have stayed at Los Alamos... but Bradbury maintained a cautious approach, then and throughout his career” – Teller

“He maintained that a thermonuclear weapon could not be built in the foreseeable future. It was clear to me that all meaningful work on the project had ended for the foreseeable future.” – Teller

Norris Bradbury and Edward Teller Ontology in Protégé

Class hierarchy:



To track significant moments in this relationship, the ontology is divided by 4 classes: major “Events”, their “Effects”, “MajorTopics”, and “Subtopics”.

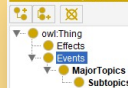
Active ontology x Entities x Classes x Object p

Object property hierarchy: owl:topObjectProperty

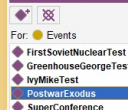


Object properties such as *IsConsequenceOf* create relationships between individuals of classes.

Class hierarchy: Events

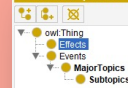


Direct instances: PostwarExodus

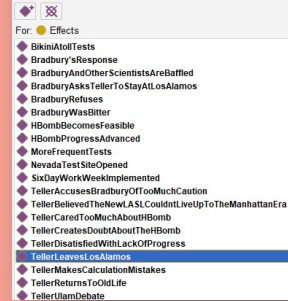


In this case, the individual is *PostwarExodus* of the class *Events*.

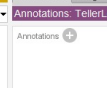
Class hierarchy: Effects



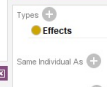
Direct instances: TellerLeavesLosAlamos



Annotations: TellerLeavesLosAlamos



Description: TellerLeavesLosAlamos

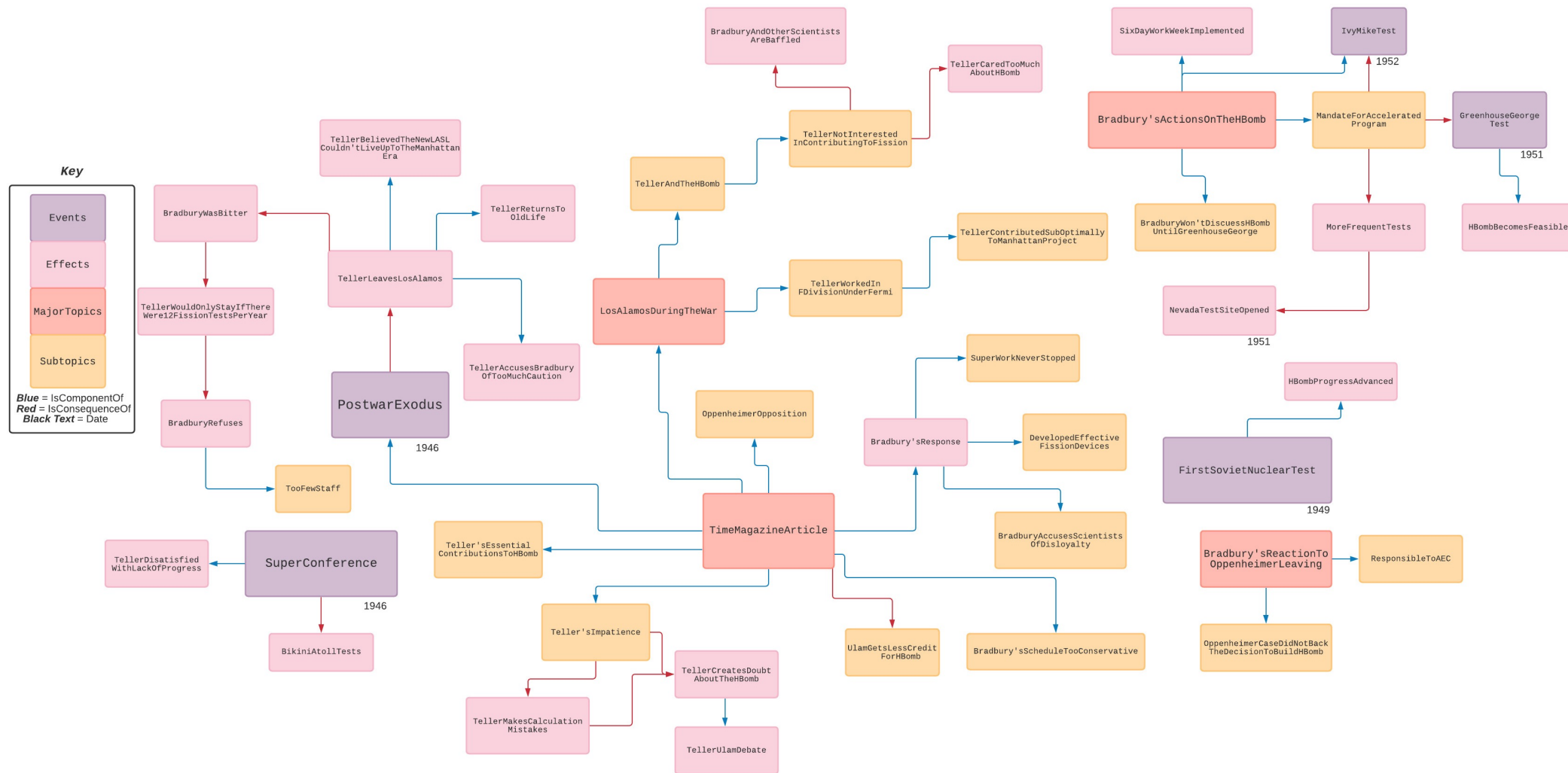


Property assertions: TellerLeavesLosAlamos



To communicate a causal relationship between the “Event” *PostwarExodus* and the “Effect” *TellerLeavesLosAlamos*, the object property *IsConsequenceOf* is asserted between the two. Data properties, which are specific to the individual, are used to assert the *Date* of the event.

Norris Bradbury and Edward Teller Ontology



This flowchart represents all the individuals, properties, and relationships in the ontology.

Titan on the Red



Goals of Titan on the Red

- *Provide an integrated search across data repositories*
- *Improve access to mission-critical archival information*
- *Ensure secure access*

Capabilities

- Machine learning based Optical Character Recognition (OCR)
- Machine learning to recognize metadata

Protégé and ontologies must be expanded and integrated into Titan to improve Titan's search results. This project demonstrates the crucial role ontologies perform in improving search technology



| CORE DATA PLATFORM



The Forthcoming Hydrogen Bomb Book

For the 75th Anniversary of the Ivy Mike test, the laboratory is writing a definitive history on the invention of the Hydrogen-Bomb.

Alan Carr, the senior laboratory historian and one of the lead authors, asked our student group to research the interactions between Norris Bradbury and Edward Teller to aid research

Protégé demonstrates our historical research and serves as a tool to aid computational database search process.

We have uncovered many documents in the NSRC



Special thanks to our mentors Alan Carr, Danny Alcazar, Nick Lewis, Madeline Whitacre, and Julie Maze.

Here is roughly what we plan to say:

1st Two Slides

William

_____ Why do we bother converting physical archives to digital ones? It's about accessibility, making documents easier to find and use, right?

First, it will be useful to go over a few basic parameters related to database searches: precision, recall and F-score. Precision is just what it sounds like: how many of the search results I get are actually relevant to what I was looking for in the first place? A very precise search function will not hand me any documents that aren't useful to me.

Recall relates to the quantity of search results - how many documents the system hands me when I search. Having a high recall is also important - it means I have access to more information from a single search.

_____ The really crucial metric here is called F-score. This means that the function's precision and recall are the same, so that every result I get for a search is related to what I was looking for. If a search function has a poor balance of precision and recall, my search will either get so few hits it's almost useless or so many unrelated results that any information I could use is lost in a sea of irrelevant documents. What we really want is for the system to have high recall and about equally high precision for it to be as usable as possible.

_____ Ok, what about ontologies? An ontology in this context is a network of discrete concepts (things, people, places, etc.) and, importantly, the relationships that connect them. A robust ontology is key to making a database searchable; if I search for something like "Edward Teller super", a search engine with an appropriate ontology will be able to draw a web of relationships among concepts like "Edward Teller," "nuclear physics," "Los Alamos," "hydrogen bomb," etc. This will help both the recall and the precision of the search: being able to link to concepts I didn't type directly into the search box will get more results from a search, while knowing how the concepts are related to each other will make it easier to filter out the stuff that isn't actually useful. Ontologies make a database more flexible and easier for a user to actually navigate.

3rd, 4th Slides

Thomas

Slide one:

We created an ontology as proof of concept and structured it around the tense relationship between Norris Bradbury and Edward Teller. Norris Bradbury was the second laboratory director and was instrumental in Los Alamos' development of the Hydrogen Bomb. Bradbury worked in

the explosives division during the Manhattan Project and was widely acknowledged to be a brilliant administrator. He succeeded Oppenheimer as director and held the position until 1970. Edward Teller was a brilliant theoretical physicist who is widely regarded as the principle originator of the Hydrogen bomb. Teller worked at Los Alamos on issues related to thermonuclear weapons, and came back intermittently to consult for Los Alamos in the immediate postwar years.

These two men had different personalities and different strengths. Bradbury was a great administrator, a man who was practical. Teller, in contrast, was a great theoretician, but he lacked practical engineering expertise. Teller was also known to be susceptible to paranoia. These two men had a turbulent relationship over their differences regarding the development of the H-Bomb.

Slide Two:

During the period between 1946 to 1949, the lab's central focus was to make fission bombs more efficient and more compact. Teller believed that Bradbury took a too-conservative approach to the development of thermonuclear weapons, and advocated putting more resources towards the H-Bomb. Contrary to Teller's accusations, there were not three lost years in the development of Hydrogen Bombs from 46-49, and the work to make the fission bombs more compact were crucial to the development of the success of Mike and later test series. The quotes on this slide document this conflict in the words of Bradbury and Teller. Bradbury was frustrated by Teller's obsession with thermonuclear weapons and his inability to focus on broader and interrelated challenges. Teller charged Bradbury with excessive caution. Teller ultimately deserves substantial credit for his theoretical work on the H-Bomb, but Bradbury's approach should be lauded. LASL successfully developed hydrogen bombs.

5th, 6th Slides

Sydney

Slide one:

- To organize the complex relationship between Bradbury and Teller, we've used a program called Protégé to build an ontology based on it. This is how the ontology appears in Protégé...
- To track the significant moments in Bradbury and Teller's relationship, our ontology has been divided into four defining classes: major Events, their Effects, MajorTopics, and Subtopics, all of which had a part in shaping their relationship.
- Object properties are used to create relationships between individuals within the previously mentioned classes. For example, "IsConsequenceOf" shows that one individual caused another. In this case, our individual is PostwarExodus of the class Events.

- To communicate a relationship between the Event Postwar Exodus and the Effect TellerLeavesLosAlamos, the IsConsequenceOf object property is asserted between them. Ultimately, this creates the idea that Teller leaving Los Alamos was a consequence of the Postwar Exodus.
- Furthermore, data properties are used to specify characteristics of certain individuals and not create relationships between them, like object properties. In this ontology, we've created the data property Date. This allows us to assert that Teller left in 1946.

Slide two:

- This flowchart is representative of all the individuals and relationships in our ontology. As you can see, purple rectangles represent individuals of the Event class, pink is Effects, red is MajorTopics, and yellow is Subtopics.
- These are connected by lines, which represent object properties - blue being IsComponentOf, red being IsConsequenceOf (refer to PostwarExodus example). Any black text beneath the rectangles represents the data property Date.
- This basic model is just a taste of how complex relationships and topics can be easily organized by utilizing ontologies.

7th, 8th slides

Camille

Slide One:

- Titan on the Red is an artificial intelligence system that allows LANL to automatically catalog documents and make its search engines/archives more effective and accessible.
- LANL has over 10 million pieces of media in the NSRC - only 10% of which is accessible/cataloged. Further, manually cataloging all of this would take more than 400 years. The documents that are readily accessible with a search query are only the tip of the iceberg and are very easily reachable.
- Titan, as well as the ontologies integrated within it, would help LANL automatically catalog the massive amounts of documents that are yet to be digitized, as well as make all of them readily available with a "google-like" search engine.
- When a document is scanned, Titan creates an OCR (Optical Character Recognition) output. Meaning it reads the text and digitizes it automatically. This automatic process eliminates the need for people to manually type out documents that need to be cataloged.
- Protege is an ontology-building tool that can be integrated into Titan in order to aid search results. We hope that our demonstration of protege highlights the importance of integrating ontologies into our search tools.

Slide 2:

- Thomas spoke about the relationship between Edward Teller and Norris Bradbury.

- This topic was selected by our laboratory historians to aid their research for the forthcoming book by LANL historians. There is a new effort underway to write a definitive history of the hydrogen-bomb development at Los Alamos for the 75th Anniversary of the Ivy Mike test.
- Bradbury and Teller will be important characters in the forthcoming book, and their relationship as well as their work are crucial to understand the story of the H-Bomb. Thus, Protege represents some of the research that we conducted and also can be used as a tool to facilitate computational database search processes.
- Furthermore, many of the documents we utilized are from the NSRC.